

Y-nvestigation: applicazione per analisi geografica della variabilità del cromosoma Y

Davide Petricone[^], Andrea Novelletto[^], Loredana Vigliano^{*}

[^]Dipartimento di Biologia, Università Tor Vergata Roma, petrico@hotmail.com

[^]Dipartimento di Biologia, Università Tor Vergata Roma, novelletto@bio.uniroma2.it

^{*}Dipartimento di Ingegneria dell'Impresa, Università Tor Vergata Roma, vigliano@mat.uniroma2.it

Riassunto

In questi ultimi anni è stato ampiamente dimostrato e riportato dalla letteratura scientifica che, per scopi medico-legali e antropologici, vi è molto spesso la necessità di utilizzare anche dati di DNA mitocondriale e del NRY (regione non ricombinante sul cromosoma Y) in grado di combinare le informazioni sul genoma con informazioni geografiche.

La somma di queste due informazioni consente di individuare le regioni nel mondo dove la frequenza di marcatori, o combinazioni di marcatori identici a quelli cercati (ad esempio in una traccia biologica qualsiasi) è più alta, ovvero di attribuire una o più regioni più probabili come origini del soggetto o della traccia.

In particolare i marcatori del cromosoma Y ereditati con la linea paterna consentono la ricostruzione della storia naturale sulla base delle mutazioni via via acquisite.

L'identificazione di sotto-aplogruppi (che fanno riferimento alla variazione della regione non ricombinante del cromosoma Y), la definizione della loro distribuzione geografica e la quantificazione della loro variazione interna, possono dare importanti informazioni sui processi migratori antichi e recenti e su eventi demografici che, avendo lasciato tracce nella struttura genetica delle popolazioni moderne, ce ne raccontano la storia.

L'applicazione web sviluppata, chiamata Y-nvestigation, permette, tramite la costruzione di apposite banche dati, di raccogliere informazioni di tipo biologico su tale cromosoma, combinare tali informazioni sul genoma con i relativi dati geografici, esplorare quindi la variazione genetica e la distribuzione geografica dei tipi di cromosomi Y tramite la visualizzazione di mappe principalmente dello spazio eurasiatico.

Introduzione ed obiettivi

La maggioranza delle informazioni che ogni individuo maschile porta sul proprio cromosoma Y sono ereditate per via patrilineare e rappresentano la somma delle mutazioni avvenute nel corso dell'evoluzione durante le generazioni che si sono susseguite. Tali informazioni si trovano sulla regione non ricombinante del cromosoma Y denominata NRY (Non-Recombining Y-chromosome) e sono, nello specifico, dei marcatori polimorfici utilizzati proprio per la

caratterizzazione del cromosoma stesso. I marcatori sono polimorfismi microsatelliti o STR (Y-STR, short tandem repeat) e i polimorfismi a singolo nucleotide SNP (single nucleotide polymorphism).

In definitiva, nella regione ricombinante del cromosoma Y (NRY) di ogni individuo maschio si possono trovare la somma delle mutazioni evolutive di migliaia e migliaia di anni che racchiudono anche informazioni geografiche relative ai luoghi ove i suddetti marker si sono originati e le vie attraverso le quali sono stati dispersi.

Il nostro primo lavoro, a titolo sperimentale, è stato quello di raccogliere queste informazioni geografiche relative ai marker di soggetti appartenenti alle popolazioni eurasiatiche, direttamente dagli articoli pubblicati, e farne una banca dati. Tutti i dati risultano scientificamente accertati.

La combinazione delle tipizzazioni genetiche con le informazioni geografiche consente di individuare le regioni nel mondo dove la frequenza di marcatori, o combinazioni di marcatori, è più alta, ovvero di attribuire una o più regioni più probabili come origini del cromosoma Y del soggetto o della traccia.

Abbiamo quindi sviluppato un'applicazione web per l'interrogazione di tale base di dati, con l'obiettivo specifico di consentire l'analisi geografica della variabilità spaziale del cromosoma Y al fine di ripercorrere il tracciato territoriale o geografico a ritroso del «cammino percorso dai nostri avi».

Cromosoma Y e terminologia.

L' Y è tra i cromosomi più piccoli del nostro genoma, è aploide e non contiene informazioni essenziali per la vita di un individuo. Esso è trasmesso unicamente da un padre a tutti i suoi figli maschi. La gran parte dei geni sono localizzati nella regione eucromatica (104 geni), mentre la restante parte del suo DNA va a formare un'ampia regione eterocromatica.

D'interesse in questo lavoro è il sottoinsieme di marcatori situati sulla cosiddetta regione non ricombinante sull'Y. Con l'azzeramento del tasso di ricombinazione, tale regione ha mantenuto particolari combinazioni di mutazioni.

Il numero di marcatori polimorfici all'interno è grande, e la loro combinazione in cis (aplotipo) può raggiungere un elevato potere discriminante tra cromosomi. I marcatori sono:

1. polimorfismi a singolo nucleotide, o SNP (single nucleotide polymorphism).
2. polimorfismi microsatelliti o STR (Y-STR, short tandem repeat)

Ciò che è per noi interessante è che tutti questi marcatori, sono trasmessi in maniera esattamente identica da padre in figlio maschio, cioè nella stessa combinazione di alleli (aplotipo). Con le generazioni successive gli aplotipi possono accumulare nuove variazioni soltanto per mutazione.

Sulla base dei particolari SNPs osservati è stato costruito un albero filogenetico i cui rami (Aplotruppi) sono indicati con le lettere da A a T. Per ciascun ramo può essere indicata la particolare mutazione che lo identifica.

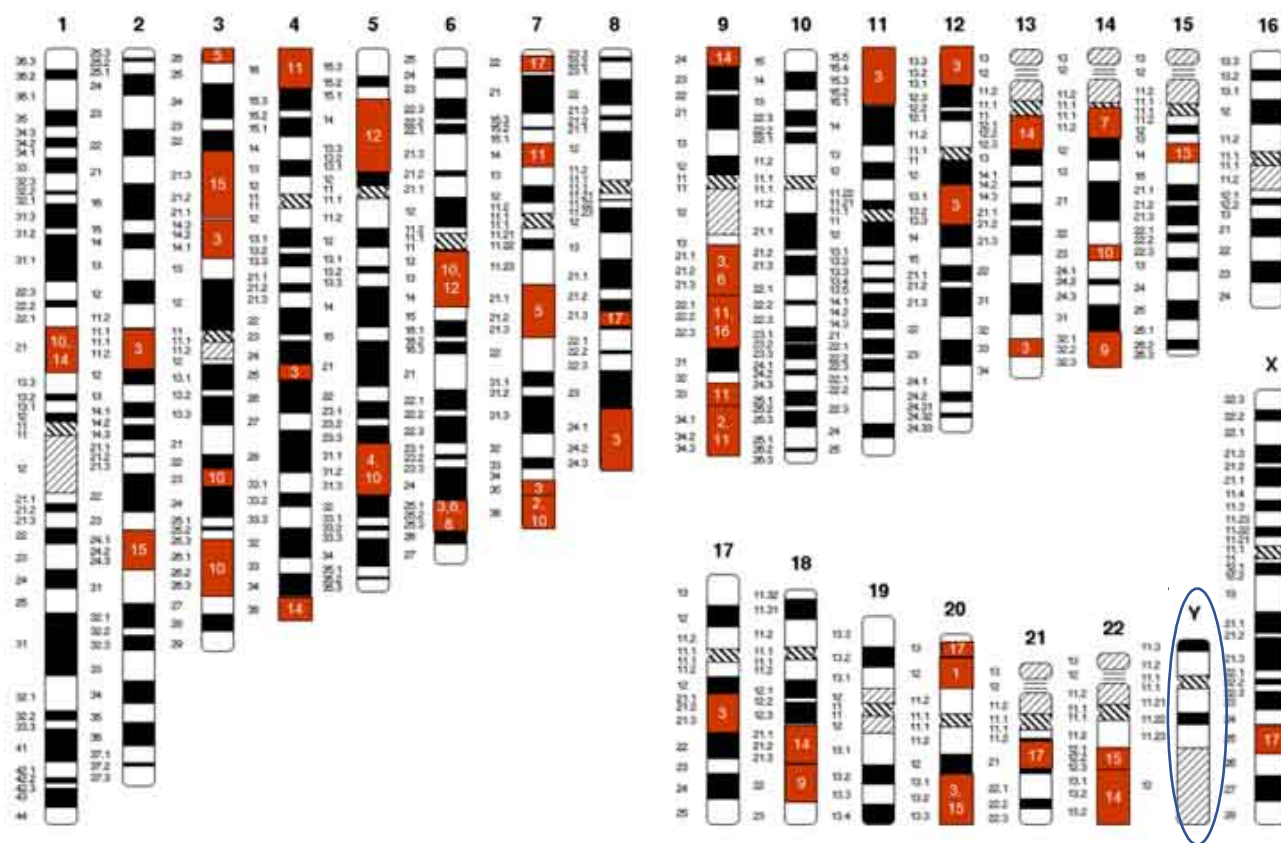


Figura 1: Mappatura del genoma umano con tutti i cromosomi, compreso il cromosoma Y, i colori e i numeri identificano rispettivamente le diverse regioni del DNA (codificante, non codificante) e le diverse regioni di ogni singolo cromosoma.

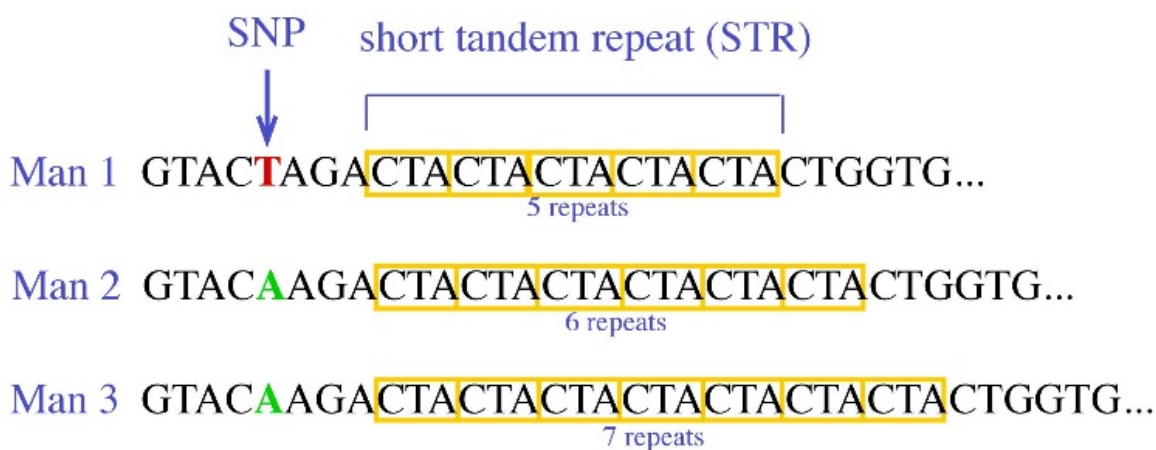


Figura 2: STR e SNP

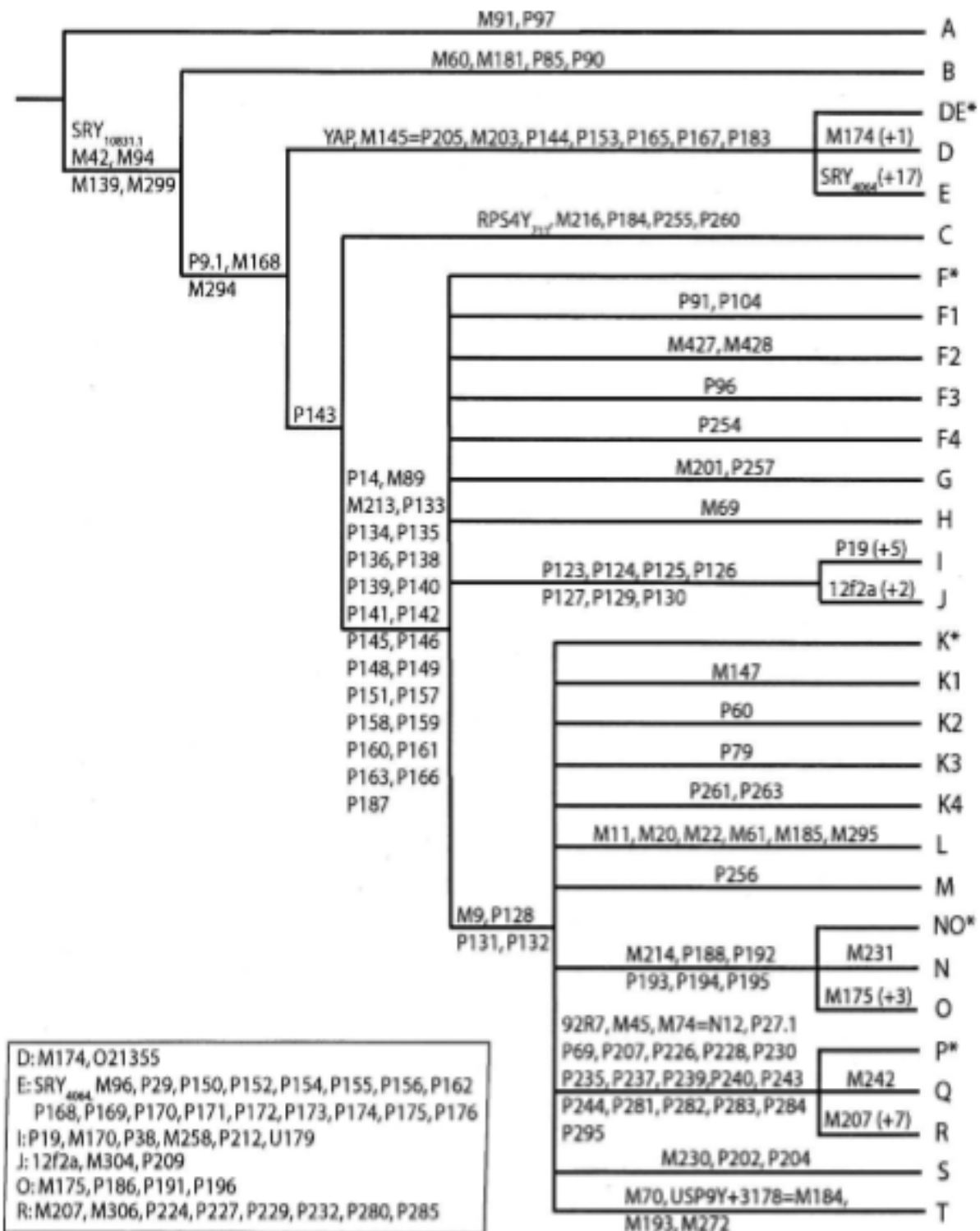


Figura 3: Una forma abbreviata dell'albero della parsimonia del cromosoma Y. I nomi delle mutazioni sono indicati sui rami. I sottoalberi corrispondenti ai cladi maggiori A-T sono compressi in questa figura. I nomi degli aplogruppi sono dati in cima all'albero, e i cladi principali sono etichettati con lettere maiuscole grandi. I nomi delle mutazioni sono dati lungo i rami; la lunghezza di ciascun ramo non è proporzionale al numero di mutazioni o all'età della mutazione. L'ordine dei marker filogeneticamente equivalenti mostrati su ciascun ramo è arbitrario. (Karafet et al. 2008).

Per quello che riguarda lo scopo del nostro lavoro, molte ricerche hanno dimostrato che i marcatori dell'NRY mostrano la più alta quota di variabilità tra le popolazioni umane (Hammer et al., 2001,2003; Romualdi et al., 2002).

Ciò suggerisce che la diversità dell'NRY possa essere utilizzata per rilevare lo strutturamento popolazionistico anche su una scala microgeografica, permettendo l'analisi delle tracce genetiche di eventi quali le migrazioni, le mescolanze ed effetti del fondatore e della deriva su aree locali.

Si ha quindi la necessità di utilizzare banche dati per raccogliere queste informazioni per combinarle con informazioni geografiche.

L' applicazione Y-vestigation

L'applicazione web sviluppata in questo lavoro, chiamata Y-vestigation, permette l'esplorazione della variazione genetica e distribuzione geografica dei tipi di cromosomi Y nello spazio eurasiatico principalmente tramite l'ausilio di mappe. Il lavoro è stato quello di riorganizzare informazioni geografiche su un insieme di soggetti di popolazioni eurasiatiche tipizzati per i marcatori SNP e STR riportati in letteratura, e poi costruire e implementare un database geografico sperimentale.

Il Database

La costruzione della base di dati si è basata sulla raccolta e l'analisi di diversi articoli scientifici e nel dettaglio sono stati estrapolati:

- i dati sui luoghi geografici di campionamento dei soggetti;
- i valori degli YSTR (non tutti i campioni sono stati tipizzati per tutti i possibili STR della base di dati);
- i dati sugli SNP;
- i nomi degli aplogruppi collegati agli SNP.

L'analisi su tali informazioni ha portato alla progettazione di uno Schema Entity-Relationship del database e alla sua implementazione in MySQL.

L'analisi su tali informazioni ha portato alla progettazione di uno Schema Entity-Relationship del database e alla sua implementazione in MySQL.

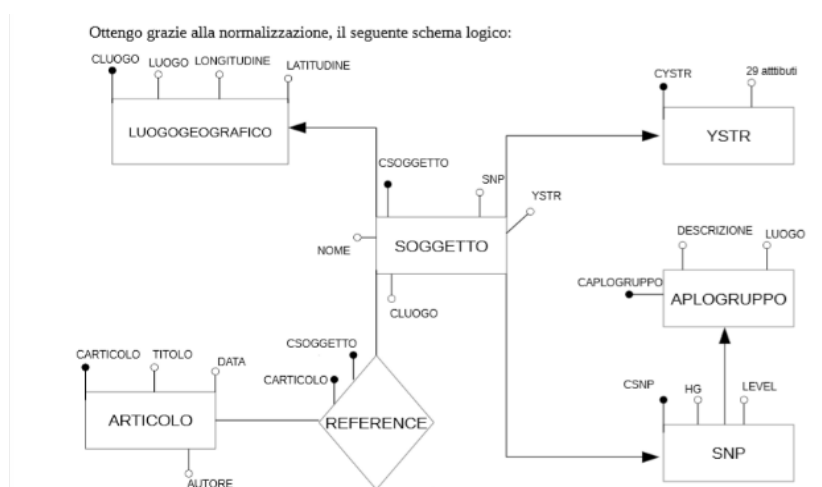


Figura 4: Schema ER normalizzato del database

Applicazione Y-nvestigation

E' stata quindi sviluppata poi un'applicazione web per interrogare il database, con lo scopo specifico di riassumere le informazioni geografiche della variabilità spaziale della NRY e con l'obiettivo ultimo di poter percorrere a ritroso il cammino dei maschi attraverso cui determinati aplogruppi sono giunti fino nei luoghi dove li troviamo oggi, in pratica ripercorrere il cammino dei nostri avi. Il sito web, che ora è visualizzato solo sul server locale, è in grado di raccogliere informazioni di tipo biologico su tale cromosoma, combinare tali informazioni sul genoma con i relativi dati geografici, esplorare infine la variazione genetica e la distribuzione geografica dei tipi di cromosomi Y tramite la visualizzazione di mappe.

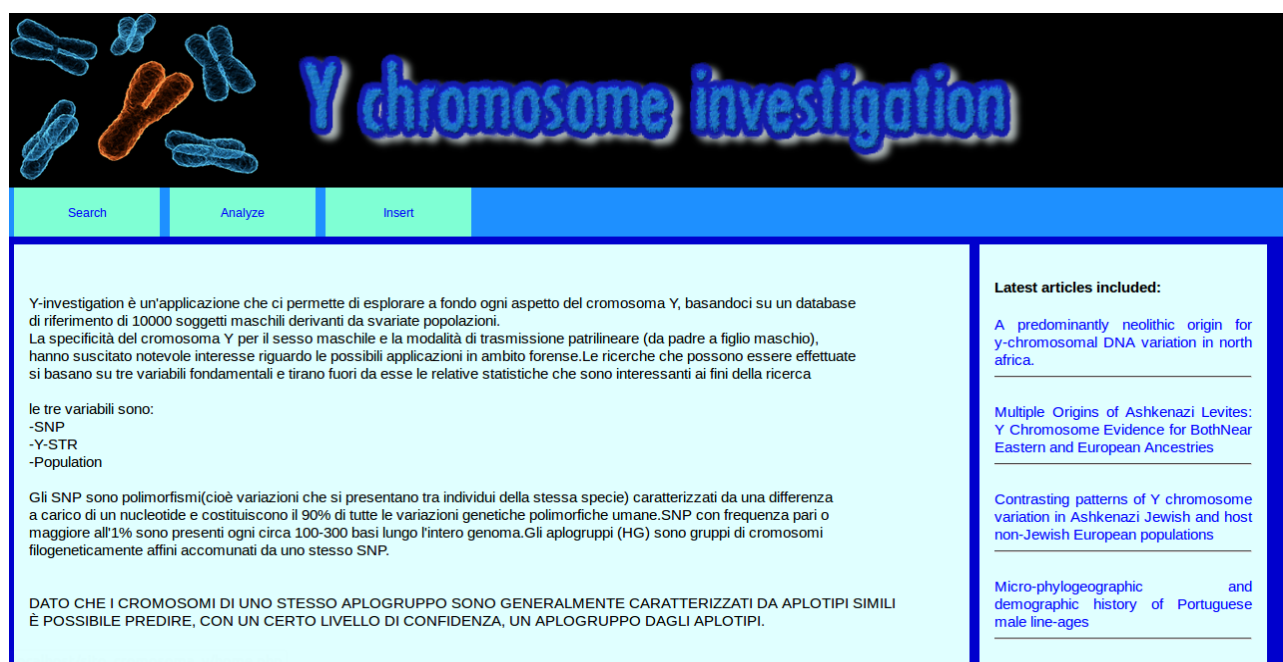


Figura 5: Y chromosome investigation

L'applicazione può inoltre interrogare il database in tre modi diversi:

- YSTR(aplotipi): Scegliendo uno specifico aplotipo, da menu a tendina, vengono elencate le rispettive informazioni e statistiche sugli SNPs e Popolazioni in cui vi è presenza di quel determinato assetto genetico.
- SNPs: Scegliendo lo SNP d'interesse andiamo a identificare le popolazioni e/o aplotipi che condividono questo specifico SNP.
- POPULATION: Scegliendo una determinata popolazione o localizzazione geografica, è possibile mostrare la distribuzione di tutti gli assetti aplotipici presenti in quella regione.



NUMBER OF SUBJECTS:	PLACE:	PERCENTAGE OF THE POPULATION
1	Ethiopia C	6.67
1	Czech Rep.	3.23
1	Italy Sicily	0.93
1	Germany C	6.25
1	India C	20
1	Portugal South	1.92
1	France Bearnais	50
1	Turkey region 6	1.69
1	Russia European North	0.82
1	Tunisia	0.73
1	Ukraine	2.38
1	Greece Chios	14.29
1	Italy Campania	1

Figura 6: Esempio di query per uno specifico Aplotipo mediterraneo, la tabella rappresenta la percentuale di presenza di quello specifico aplotipo in Europa (calcolata sulla base dei dati presenti nel nostro database).

Il sito comprende, tra le altre, una sezione aggiuntiva in cui tutti i criteri di ricerca possono essere combinati, con altri campi significativi, ad esempio con:

- **GEOGRAPHICAL WINDOW:** Questo campo ha una selezione obbligatoria qualora venga indicata una determinata popolazione nel campo POPULATION, e serve per espandere la finestra geografica intorno alla popolazione prescelta, sia in longitudine che in latitudine, di 1,2 o 3 gradi.

Un caso di studio : la Sardegna

Come abbiamo sottolineato, i marcatori del cromosoma Y ereditati nella linea paterna consentono la ricostruzione della storia naturale sulla base delle mutazioni via via acquisite. Concentreremo ora il nostro studio sulla popolazione umana sarda andando ad usufruire di tutte le potenzialità dell'applicazione web Y-nvestigation con lo scopo di dare una più accurata analisi genetico-geografica basata su tale variabilità in questa popolazione.

Una prima ricerca che inizia a dare i suoi più evidenti risultati è proprio quella che parte dal nome della popolazione e, quindi, nella sezione Population abbiamo scelto dal menu a tendina la voce "Italy Sardinia S", ovvero sud della Sardegna, dove abbiamo a nostra disposizione nella banca dati l'informazione su 242 soggetti. Dalla ricerca effettuata, sono due i risultati importanti che catturano subito la nostra attenzione e sono quelli che poi maggiormente racchiudono l'informazione generale sulla variabilità del cromosoma Y. In primo luogo ci viene mostrata la presenza di vari marcatori SNP all'interno della popolazione e questi all'interno del proprio cluster hanno diverse sequenze aploipiche (STR). In secondo luogo notiamo di conseguenza tre SNP (Figura 7) in questa popolazione che vantano una percentuale maggiore rispetto agli altri: M26,M201,M269.

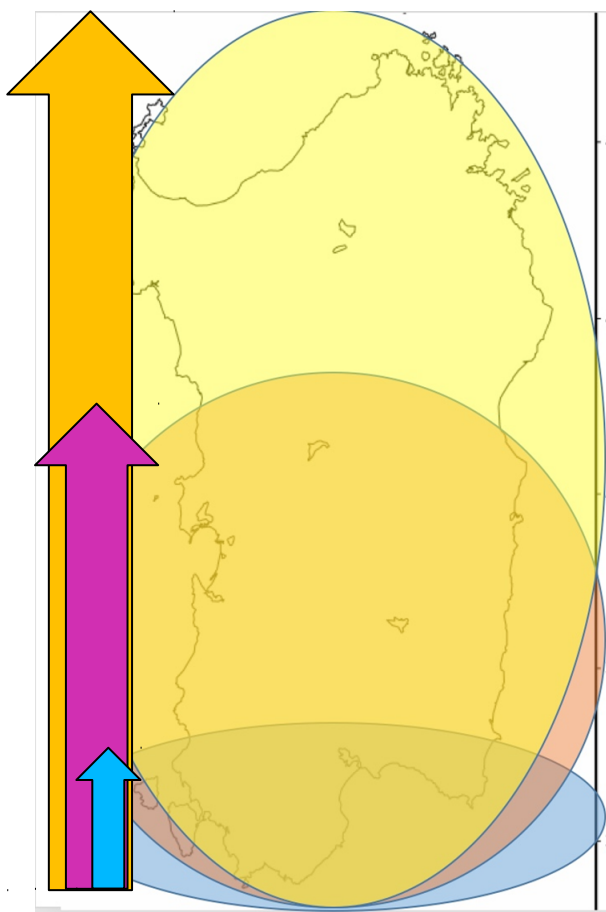


Figura 7: Gli SNP M26,M201, M269 prevalenti al sud della Sardegna

Il confronto di un qualsiasi SNP o STR anche in una piccola realtà geografica e l'area immediatamente adiacente più grande è di enorme utilità per la comprensione delle dinamiche migratorie delle popolazioni umane. La finestra geografica su tutta la Sardegna, ci ha portato a concludere che gli SNP più presenti in questa regione sono sempre gli stessi del sud della Sardegna, ma questa volta con una percentuale più alta per lo SNP M26. Ciò mostra un possibile arricchimento di questo aplotipo man mano che ci si muove dal sud al nord della Sardegna. Alla base di questa osservazione potrebbe esservi una direzionalità (da sud a nord) preferenziale delle migrazioni umane antiche anche in una piccola realtà geografica come quella della Sardegna (oppure una concentrazione e una crescita demografica maggiore dei portatori di M26 nel nord).

Conclusioni e sviluppi futuri

L'applicazione web sviluppata in questo lavoro, chiamata Y-nvestigation, permette l'esplorazione della variazione genetica e distribuzione geografica dei tipi di cromosomi Y nello spazio eurasiatico soprattutto con l'ausilio di mappe. La sua utilità ci è stata dimostrata con l'applicazione al caso di studio della Sardegna, che ha mostrato come, anche in un'area geografica ristretta, sia possibile ricostruire le tracce dei cambiamenti demografici e biogeografici dei nostri avi.

Gli obiettivi futuri saranno quelli di:

- ampliare la base di dati per arrivare a mantenerla continuamente aggiornata sulle differenti combinazioni aplotipiche riscontrate nella letteratura e in determinate zone geografiche con i relativi SNP che differenziano i diversi aplogruppi. Ciò renderà possibile affrontare indagini sempre più complesse, anche a livello microgeografico;
- creare ulteriori cartogrammi di frequenze relativi alla distribuzione che nel tempo i vari marker hanno assunto nelle diverse zone, anche in realtà geografiche più piccole, per arrivare a comprendere il percorso che hanno compiuto le diverse popolazioni nel corso della storia.

Bibliografia

- Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., Demaine, A.G., Barbujani, G., Previderè, C., Wilson, I.J., Tyler-Smith, C. and Jobling, M.A. (2010). A predominantly neolithic origin for European paternal lineages. *PLoS Biol* 19,8(1):e1000285.
- Contu, D., Morelli, L., Santoni, F., Foster, J.W., Francalacci, P. and Cucca, F. (2008). Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: inference for association scans. *PLoS One*. 9,3(1):e1430.
- Elmasri R., Navathe S. Sistemi di basi di dati, Fondamenti e complementi, Pearson 7a edizione italiana

Google Maps API <https://cloud.google.com/maps-platform/>.

GRASS Development Team (2016). Geographic Resources Analysis Support System (GRASS) Software, Version 7.03. Open Source Geospatial Foundation. <http://grass.osgeo.org>.

Graves, J.A., Koina, E., and Sankovic, N. (2006). How the gene content of human sex chromosomes evolved. *Curr. Opin. Genet. Dev.* 16(3):219-24.

Hammer, M.F., Karafet, T.M., Redd, A.J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodiyall, H., and Zegura, S.L. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18, 1189-1203

Hammer, M.F., Blackmer, F., Garrigan, D., Nachman, M.W., and Wilder, J.A. (2003). Human population structure and its effects on sampling Y chromosome sequence variation. *Genetics* 164, 1495-1509.

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18, 830-838.

Myres, N.M., Rootsi, S., Lin, A.A., Järve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Pshenichnov, A., Yunusbayev, B., Balanovsky, O., Balanovska, E., Rudan, P., Baldovic, M., Herrera, R.J., Chiaroni, J., Di Cristofaro, J., Villems, R., Kivisild, T., and Underhill P.A. (2011).) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* 19, 95-101

Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A., and Barbujani, G. (2002). Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 12, 602-612.



©2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).